DOCUMENT RESUME

ED 243 916 TM 840 212

AUTHOR Benson, Jeri; Zarnegar, Zohreh
TITLE Testing the Structural and Theoretical Aspects of Construct Validity Using Confirmatory Factor Analysis.
PUB DATE Apr 84
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Education; Factor Analysis; *Factor Structure; *Goodness of Fit; *Models; Multidimensional Scaling; Research Methodology; *Self Concept Measures; *Test Validity
IDENTIFIERS Aggregation (Data); *Confirmatory Factor Analysis; LISREL Computer Program; *Piers Harris Childrens Self Concept Scale

ABSTRACT

The purpose of the study was to use confirmatory factor analysis techniques to assess the adequacy of previous validity studies related to the structural and theoretical dimensions underlying a measure of self-concept, the Piers Harris Children's Self Concept Scale. Responses from 890 students in grades 3-6 were used as input to the LISREL program to test several theoretical models thought to represent the data. In addition, three subsamples were also created and used to test the effect of aggregating data on the structural validity of the scale by testing a given factor structure within subgroups simultaneously. Results indicated that a multidimensional model fit the data better than a general one-factor model. However, the multidimensional model was not invariant across the subgroups. Confirmatory factor analysis in construct validity studies is thought to be superior to either exploratory factor analysis or multitrait-multimethod matrices in its ability to analyze the theoretical and structural dimensions of a scale. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243916

TESTING THE STRUCTURAL AND THEORETICAL ASPECTS OF CONSTRUCT VALIDITY USING CONFIRMATORY FACTOR ANALYSIS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Jeri Benson

University of Maryland

Zohreh Zarnegar

University of Southern California

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Benson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

The purpose of the study was to use confirmatory factor analysis techniques to assess the adequacy of previous validity studies related to the structural and theoretical dimensions underlying a measure of self-concept. Responses from 890 students in grades 3-6 were used as input to the LISREL program to test several theoretical models thought to represent the data. In addition, three subsamples were also created and used to test the effect of aggregating data on the structural validity of the scale by testing a given factor structure within subgroups simultaneously. Results indicated that a multidimensional model fit the data better than a general one factor model. However, the multidimensional model was not invariant across the subgroups. Implications for using confirmatory factor analysis in construct validity studies were discussed.

TM 840 212

Presented at the annual meeting of the
American Educational Research Association
New Orleans, April 1984.

TESTING THE STRUCTURAL AND THEORETICAL ASPECTS OF CONSTRUCT VALIDITY USING CONFIRMATORY FACTOR ANALYSIS

The use of exploratory factory analysis and multitrait-mulitmethod matrix approaches to establish the construct validity of an instrument is well documented in the measurement lituratue. More recently, the use of confirmatory factor analysis developed by Joreskog and Sorbom (1983) has been shown to be a methodologically superior technique for the assessment of construct validity (Shavelson & Bolus, 1982). The purpose of the present study is to illustrate how confirmatory factor analysis can be used to evaluate conflicting research surrounding the theoretical and structural dimensions that underlie one of the more widely used measures of self-concept, the Piers-Harris Children's Self-Concept Scale (Piers & Harris, 1964).

The Piers-Harris scale has received positive reviews by numerous researchers (Bentler, 1972; Robinson & Shaver, 1973; Shavelson, Hubner & Stanton, 1976; Smith & Rogers, 1977; Wylie, 1974). However, conflicting evidence as to the nature of the factor structure underlying the Piers-Harris scale has been reported by Bentler (1972), Cowan, Altman and Pysh (1978), Lynch and Chaves (1975), Marx and Winne (1980), Michael, Smith and Michael (1975), Moran, Michael and Dembo (1978) and Platten and Williams (1979, 1981). Many of these studies had used either exploratory factor analysis or multitrait-multimethod matrices in attempting to establish the nature and number of dimensions thought to be measured by the Piers-Harris scale.

Several problems identified by Shavelson et al. (1976) plague

the previous efforts to validate the construct of self-concept ranging from imprecise definitions, the nonequivalence of various self-concept measures to interpretation problems due to the self-report techniues used. This paper addresses the issue of intrepretational problems due to traditional validation techniques. There are at least two reasons for these problems. First, the use of statistical procedures that rely on subjective criteria to evaluate the theoretical dimensions of construct validity, e.g., multitrait-multimethod and traditional factor analytic procedures. Second, the statistical procedures have been run on aggregated data combining sex, ethnicity socio-economic status and grade level assuming these variables would not affect the structure of the construct being measured.

Multitrait-multimethod studies of construct validity involving the Piers-Harris scale (Franklin, Duley, Rousseau & Marsh & Smith, 1982; Marx & Winne, 1978; Winne, Marx & Taylor, 1977) are often limited by subjective criteria such as how high should the convergent correlations be and how low should the discriminant correlations be, what methods and traits should be included and problems of method bias and unreliability of measures. In factor analytic studies of the Piers-Harris (Platten & Willliams, 1979, 1981; Michael et al., 1975; Moran et al., 1978) the problems become which factor analytic and rotation solution to use and how may factors to rotate and interpret. These problems have been discussed by Marsh and Hocevar (1983) and Hocevar and Hocevar (1982). The procedure they recommend and the one advocated here is the use of confirmatory factor analysis. Using confirmatory factor analysis techniques via

LISREL procedures (Joreskog & Sorbom, 1983) to establish the construct validity of a scale allows one to test various competing models of thought to be measured by an instrument.

The Piers-Harris scale (The Way I Feel About Myself) was designed primarily for "research on the development of children's self attitudes" (Piers, 1969, p.2). Piers and Harris developed an 80 item survey thought to assess six content factors. While their manual provides normative data for an overall self-concept score, the six content factors have frequently been analyzed and discussed in the liturature. In their article, Shavelson et al., (1976) take the position that self-concept is multidimensional and hierarchial. Much the factor analytic and the multitrait-multimethod literature on the Piers-Harris scale tend to support the Shavelson et al. contention (Marsh & Smith, 1982; Michael et al., 1975; Moran et al., 1978; Shavelson & Bolus, 1982). However, as Marsh and Smith (1982) have pointed out, the specific dimensions obtained from these analyses are not always clear nor consistent across samples.

A different line of reasoning is presented by Bentler (1972). In his review of the Piers-Harris scale, he indicated that perhaps fewer than six factors may represent the items on the scale and suggested that a unidimensional scale may more appropriately represent the data due primarily to the binary nature of the items on the scale. Platten and Williams (1981) have suggested that the scale be viewed as undimensional after factor analyzing pre and posttest results from the same group of subjects and noting the large proportion of variance on the first

unrotated latent root and the instability of the factor structure from pre to posttesting. In conducting multitrait-multimethod assessments of the validity of several well used self-concept instruments, Winne, Marx and Taylor (1977) and Marx and Winne (1978) found no evidence of discriminant vaidity for the multiple diminsions thought to be represented by the 80 items on the Piers-Harris scale. Confusing the picture even more, recent research on the effect of item phrasing on children's responses to attitude scales indicates that alternating phrasing may confound and obscure the construct validity of an affective scale (Andrich, 1983; Benson & Hocevar, 1983). Bentler, in his review of the Piers-Harris scale also noted that item phrasing may be effecting the dimensionality of the scale. In this case, the scale may be be reflecting two factors related to item phrasing and not item content.

As illustrated by Shavelson and Bolus (1982) and Alwin and Jackson (1981), covariance structure analysis provides a far superior method to evaluate the structural and theoretical aspects of construct validity than either traditional factor analysis or the multitrait-mulitmethod technique alone. Thus, whether the Piers-Harris scale is composed of a) six content factors as reported by Piers and Harris (1964), b) two factors based on item phrasing, c) one factor as suggested by Bentler (1972), Winne, Marx and Taylor (1977) and Platten and Williams (1979, 1981) or d) some other number of factors, the hypothesized theoretical model (number and nature of dimensions) and structural model (stability of dimensions across subgroups) can be tested and evaluated using a chi-square test of model-data

fit. Therefore, the present paper is an attempt to resolve the above conflicting conclusions found in the literature regarding the dimensionality and hence interpretability of the Piers-Harris scale.

METHODOLOGY

Subjects

There were a total of 890 students in the study from grades 3-6. The students were part of a voluntary integration program in a large urban school district. The mean age of the sample was 9.66, 60 percent were males and 40 percent were females. The ethnic breakdown was 38 percent Black (n=334), 38 percent Hispanic (n=338) and 24 percent White (n=213).

Instrumentation

The Pier-Harris scale consists of 80 declarative statements to which the student responds yes or no. Thirty-six of the statements are positively phrased and 44 are negatively phrased. The items were developed from a pool of statements generated by children about what they like and disliked about themselves collected by Jersild (1952). The scale is appropriate for students in grades 3-12 and can be group administered. However, it is suggested that for grades 3-6 the items be read aloud to the group.

Items are scored one or zero based upon judges ratings of children's responses that would be indictive of high self-concept. Thus, scores range from 0 to 80 the higher score

reflecting a more positive self-concept. For comparative purposes with the data gathered in this study, the mean and standard deviation from the normative data in the Piers-harris manual was 51.84 and 13.87 respectively. The internal consistency reliabilities reported in the manual range from .78 to .93.

The constuct validity of the Piers-Harris is assessed primarily via factor analysis. Piers and Harris report a structure based upon 457 sixth grade students. Using principal components analysis with Varimax rotation, 10 factors representing 42 percent of the variance were reported. However, only six of the factors were interpreted. Sixty-seven of the 80 items loaded .30 or higher on the following six factors: Behavior (17 items), Intellectual and School Status (14 items), Physical Appearence and Attributes (9 items), Anxiety (9 items), Popularity (10 items) and Happiness and Satisfaction (8 items).

Procedure

The test was administered to groups of students that varied in size from 6-30. The items were read aloud to all children while they read along silently and then marked their response. Items were clarified when unfamiliar words appeard in the stem (e.g., I have lots of 'pep', I am 'obedient' at home) or for most of the negatively phrased items (e.g., I am unpopular). Items where the child had to disagree with a negatively phrased statement in order to give an affirmative answer was particularly confusing to almost all the students regardless of age.

The items were scored according to the protocol in the Piers-Harris manual and basic descriptive statistics obtained. The responses were then converted to covariance matrices for input into the LISREL V procedures. Six covariance matrices were formed: a) all students combined, b) one for males and one for females and c) one for each ethnic group. The following models were tested for each of the six samples:

- 1 factor model (representing one general self-concept factor
- 2 factor model (representing a possible positive/negative item phrasing effect)
- 6 factor model (representing the six content factors reported by Piers and Harris)
- n factor model (based upon a simple structure exploratory factor analysis of the data)

The models will be evaluated in terms of an overall chi-square test of fit, goodness of fit indices provided in the LISREL V output and the chi-square difference test. In addition, a critical n will be calculated to determine the sample size necessary to obtain a significant chi-square (Hoelter, 1983). Since previous research on the Piers-Harris used basically classical procedures where the errors for each item was assumed to be uncorrelated with the other items, the present confirmatory analyses will constrain the item uniquessess to be uncorrelated in order to be able to compare results. However, the factor loadings and the correlations among latent factors were set free to be estimated by the separate models tested.

The confirmatory factor analysis procedures were run in a two phase approach. First, to determine the theoretical diminsion(s) underlying the Piers-Harris scale, the most

frequently cited models found in the literature and summarized above were tested using the total sample of students. Also as part of Phase one, an exploratory factor analysis (n factor model) was run to see if a different set of factors would emerge other than that reported in the literature. The reason for this analysis was that the present sample was more ethnically diverse than samples found in the literature and this may alter the construct being measured. In Phase two of the analysis, the stability of the obtained structure obtained from Phase one was tested to determine the effect of aggregating data upon the construct of self-concept. The five separate subgroups (male, female, Black, Hispanic and White) were used. For example, the one factor model was tested simultaneoulsy across the subgroups of Black, White and Hispanic students to see if the structure fit equally well for the subgroups as it did for the total sample. This analysis was conducted for each of the theoretical models found in the literature plus the model(s) generated from the exploratory analysis in Phase one. By testing the fit of each model for each sample, it will be possible to determine whether self-concept as measured by the Piers-Harris is influenced by aggregating data by grade, gender or ethnicity.

RESULTS AND DISCUSSION

Descriptive Statistics

The mean and standard deviation for all students on the Piers-Harris scale was 57.13 and 11.65 respectively. This is slightly higher than the normative values reported in the manual.

The means, standard deviations, sample size and reliability estimates for the separate subgroups are reported in Table 1.

Table 1
Means, Standard Deviations, Sample Size and
Alpha Reliability Estimates for Subgroups

Subgroup	Mean	Standard Deviation	N	Alpha
Grade 3	58.60	10.83	240	.92
Grade 4	57.70	11.26	231	.88
Grade 5	56.36	11.54	261	.91
Grade 6	55.32	13.24	158	.92
Males	57.60	11.62	538	.91
Females	56.52	11.67	345	.92
Blacks	59.45	10.95	334	.91
Hispanics	55.22	11.49	338	.91
Whites	56.38	12.38	213	.91

The very high internal consistency estimates provide an initial indication that there is systematic covariation among the 80 items on the Piers-Harris scale.

Mean Differences. Mean differences were tested for grade level, gender and ethnicity and compared to results reported in the test manual. There were no significant differences between the scores of males and females, but there were differences among the grades and ethnic groups. The oneway ANOVA among grades ($F(3,886)=3.134$, $p<.05$) indicated that the third grade mean was greater than the sixth grade mean. This finding was similar to the pattern reported in the manual in that higher grades sometimes showed lower self-concept scores than lower grades. Piers and Harris reported that this was not a consistent pattern

and that it occurred in only a few samples. In fact, only .7 percent of variance was accounted for by grade level (omega squared). The oneway ANOVA for ethnicity ($F(2,882)=11.861$, $p<.05$) revealed that the Black students had higher self-concept scores than the White and Hispanic students. In this analysis, two percent of the variance was accounted for by ethnicity. Based upon the above results and the fact that ethnic diversity was not included in the original validation, the structural aspects of the Piers-Harris scale will be assessed only for the effect of ethnicity.

Confirmatory Analyses: Theoretical Dimensions

The results of the LISREL analyses in terms of the chi-square values and fit indices for the 1, 2 and 6 factor models based upon the literature are reported in the top half of Table 2. Two single factor models were proposed, one based upon all 80 the other using the 67 items Piers and Harris report loading on their 6 factor model. The two factor model was composed of the 36 positively phrased items set to load on one factor and the 44 negatively phrased items loading on the other factor. The six factor model is the one proposed by Pier and Harris. All the chi-square values shown in Table 2 were statiscally significant ($p<.05$), meaning that the proposed models did not fit the data well. This is not surprising since the sample size and number of variables studied is quite large. In order to get a better picture of the model-data fit, additional indices of fit are reported. While statistical criteria can help to determine the

best model given the data, statistical criteria in the absence of rationale criteria based upon the interpretability of the models proposed is essential (Joreskog & Sorbom, 1983; Marsh & Hocevar, 1984).

A commonly reported fit index is the chi-square ratio (CS Ratio - chi-square/df). There is not alot of agreement on acceptable values for the chi-square ratio. Wheaton, Muthen, Alwin and Summers (1977) suggest a ratio of 5 is acceptable, whereas Carmines and McIver (1981) suggest the ratio should be between 2 and 3. One reason for this variability in acceptable values is that the chi-square ratio like the chi-square statistic itself is sample dependent, the degrees of freedom do not reflect the sample size. For this study, chi-square ratios between 2.00 and 3.00 will be used to indicate acceptable fit. A fit statistic that is independent of sample size is the adjusted goodness of fit index (Adj GFI) obtained from the LISREL V output. The value indicates the percent of variance and covariance jointly explained by the model. Values can range from zero to one, and since there is no standard value from which to judge the present results the indices are merely descriptive and can be compared relative to the other values. Joreskog and Sorbom (1983, p. I.40) state that the goodness of fit indices are robud relative to the other values. Joreskog and Sorbom (1983, p. I.40) state that the goodness of fit indices are robust for departures from normality, which may be important considering the binary data used in the present study. Finally, critical n's (CN) have been calculated for each observed chi-square. This statistic suggested by Hoelter (1983) indicates the

sample size necessary to achieve a statistically acceptable fit using a probability level of .05. The value of critical n is calculated from:

$$CN = \frac{[1.65 + (2(df) - 1)^{1/2}]^2}{2(\text{chi-square})/N-G} + G$$

where N is the sample size and G is the number of groups in which the model is estimated. Values of 200 or larger represent an adequate fit.

Table 2

Theoretical Models Tested Using All Subjects
N=890

Model	Number of Items	Chi-Square	df	CS Ratio	Adj GFI	CN
Models in the Literature:						
1) 1 factor	80	8998	3080	2.92	.71	322
2) 2 factor	80	7740	3079	2.51	.76	379
3) 1 factor	67	6833	2144	3.18	.73	301
4) 6 factor	67	5390	2129	2.53	.78	374
Revised Models:						
5) 6 factor	56	4515	1469	3.07	.83	313
6) 4 factor	60	5390	1740	3.16	.82	307

In order to test the goodness of fit for the models in the top half of Table 2, the chi-square difference test was used. The difference test is as follows:

$$X^2_1 - X^2_2 \quad \text{and} \quad df_1 - df_2,$$

where X^2_1 represents a more restrictive model than X^2_2, which represents an alternative model with at least one additional unconstrained parameter than in X^2_1. The difference chi-square is compared to the difference between their respective degrees of freedom. The difference chi-square tests whether a more general model (less restrictive model) fits the observed covariance matrix better than a model with one or more constraints placed on it. The chi-square difference test is only applicable for nested models obtained from the same covariance matrix. Given this limitation, only models 1 and 2 and 3 and 4 from Table 2 can be tested (model 4 is nested within 3 and model 2 is nested within 1). The X^2_{1-2} was 1258, with 1 degree of freedom and the X^2_{3-4} was 1443 with 15 degrees of freedom, both statistically significant at $p<.05$. The difference chi-square indicated that the more restrictive models 1 and 3 (where it was hypothesized that a one factor model was represented by either 80 or 67 items) did not indicate as good a fit to the original covariance matrix as did models 2 and 4.

Unfortunately, it was not possible to test the difference in fit between models 2 and 4 as they were not nested. However, the fit indices and the construct represented by model 4 over model 2, show model 4 to fit the original covariance matrix 'better' than model 2. While the criteria used to select model 4 versus model 2 are subjective, the substantative apects of model 4 over model 2 make further analysis of model 4 of more interest. This is especially true since model 2 (two factors based only upon item phrasing) did not have content factors associated with it.

It was therefore impossible to tell whether item phrasing is actually being measured by the Piers-Harris. Thus model 2 should have been subdivided into a three factor model with one factor showing all items loading one a content factor (like model 1) plus two phrasing factors. Likewise for models 3 and 4 additional factors could have been proposed to detect an item phrasing effect for the one and six factor model using 67 items. The rationale for not testing those models was based more on monetary restrictions than on theoretical reasons. Therefore, further research such as was just illustrated above, but not tested in the present study, would be necessary to be able to diagnose the potential effects of item phrasing on the Piers-Harris scale.

While the six factor model reported by Piers-Harris appeared to be the most adequate given the other models tested, it was felt that it still did not provide a good fit to the present data. Thus, a simple structure model was developed based upon the same factoring and rotation procedures (principal components and Varimax) as Piers and Harris using the entire sample. This analysis was run because of the ethnic diversity in the present sample and that perhaps a better fitting factor structure might emerge. Only items loading .30 or higher were interpreted. The first 10 principal components accounted for 35 percent of the variance (Piers and Harris reported 42 percent). The analysis indicated that either a 4 or 6 factor model might better represent the data and thus were submitted to LISREL analyses for confirmation. Sixty of the 80 items loaded on the 4 factor model and 56 items loaded on the 6 factor model. The chi-square and

fit indices for these models are reported in the bottom half of Table 2 under Revised Models.

The revised models were neither nested nor based upon the same covariance matrix used as input (they differed by the number of items), hence they could not be tested against one another nor models 1-4. Using a visual inspection of all the indices of fit shown in Table 2, it was concluded that model 4 (6 factors composed of 67 items) generated by Piers and Harris best fit the present sample (given the six models tested) and was then submitted to further analysis in the following section.

Confirmatory Analyses: Structural Dimensions

To explore the effect aggregating data might have on the structural dimensions of a construct, the total sample was broken into three subgroups based upon ethnicity and model 4 from the previous analyses was explored in detail. Four nested models were proposed to determine if the factor structure and/or item error variances differed within and across each of the three ethnic groups. Relative to subgroup sample size, Boomsma (1982) has shown that LISREL is robust for samples greater than 200. Thus, the present subgroup samples were adequate for the following analyses.

Model 1 allowed the 67 items to load on the six factors as described by Piers and Harris, as well as allowing the factors to be correlated and the item error variances to be estimated freely for each ethnic group. This model will be referred to as the 'free' model and is shown in Table 3 as the sum of the separate

chi-square values from each ethnic group. Model 1 tested whether different factor loadings, factor correlations and item error variances were being measured by the scale for each group. Model 2 was represented as a 'tau-equivalent' model where all item/factor loadings were set to be invariant across each ethnic group, but the item error variances and factor correlations were free to vary. Model 2 tested whether the factor structure was the same across the three groups. Model 3 was set to be an 'error invariant' model where the item error variances were set to be invariant across each ethnic group and was a test of whether the item error variances were similar across groups. Finally, model 4, the most restrictive model, was set to have both the item loadings and item error variances invariant. Referred to as the 'parallel' model, model 4 tested the similarity of factor structure and error variance simultaneously across the groups for the six content being measured.

Model 1 was the least restrictive, most general model and model 4 the most restrictive model. Since model 1 was nested within models 2-4 the chi-square difference test could be used to determine the model which best represented the separate covariance matrices for each ethnic group. The LISREL analyses were set so that for model 2 the factor structure for Blacks indicated that the 67 items loaded on the six content factors as reported by Piers and Harris, the factors were free to correlate and the error variances (uniquenesses) were free to be estimated. The item loadings generated from the Black sample were then held invariant for the Hispanic and White groups, but the factor correlations and error variances were free to be estimated in the

Hispanic and White groups. This procedure was followed for models 3 and 4 holding invariant the item error variances and item loadings and error variances across groups respectively. The results from the analyses is reported in Table 3.

TABLE 3
Structural Models Tested Using Subsamples
For the 6 Factor 67 Item Pier-Harris Scale

Model	Chi-Square	df	CS Ratio	Adj GFI	CN
1) Parameter Estimates Free:					
Blacks (n=334)	3,628	2129	1.70	.69	560
Hispanics (n=338)	3,723	2129	1.75	.69	560
Whites (n=213)	4,208	2129	1.98	.58	500
SUM	11,559	6387	1.81	---	509
2) Loadings Invariant: (Tau-equivalent)	11,888	6503	1.83	.65	499
3) Errors Invariant:	17,657	6503	2.72	.34	338
4) Loading/Error Inv.: (Parallel)	12,535	6637	1.89	.65	491
Model Comparisons					
Model 2 vs. Model 1	329	116	2.84		
Model 3 vs. Model 1	6098	116	52.56		
Model 4 vs. Model 1	976	250	3.90		

The chi-square difference tests shown under model comparisons in Table 3 (all significant at $p<.05$) indicated that the model with the least restrictions (model 1) was a better fit to the data than models 2-4 which imposed one or more restrictions on the data. Thus, the factor structure and item errors do vary within each ethnic group, a little more so for the Whites than the Hispanics or Blacks. The implication is that the meaning and

intrepretation given to the six content factors may not actually be the same for each subgroup of students. Substantative inspection of the factor loadings and the respective labels given to each factor within each ethnic group should be undertaken as a follow up to these results.

Since the adjusted goodness of fit index shows that 58-69 percent of the joint variance-covariance matrix of the data has been explained for the 'free' model for the subgroups, a search for a better fitting model for each ethnic group seems necessary. The critical n of 500 or more for the 'free' model indicates that the current sample size is large enough to detect a significant chi-square. Therefore, once separate accepatable models for each ethnic group is obtained, chi-square difference tests for nested models like the ones illustrated above to test for invariance would be in order. An appropriate model would be one that could explain a maximum amount of the original covariance matrix and have generally small chi-square values. Guides for determining the appropriate model for each group may be found by studying the modification indices and the residual matrices provided by LISREL to locate parameters to be freed or constrained. In addition, Joreskog and Sorbom (1983) caution that total reliance on statistical criteria to fit a model may not be the best approach as the model will have to have some substantative as well as logical meaning for interpretation. Therefore, a combination of statistical and theoretical knowldege of the nature of the construct under investigation is required for appropriate model development and testing.

Finally, further research on the Piers-Harris might be directed at 1) exploring the effect of item phrasing combined with some number of content factors and 2) testing whether the construct of self-concept is hierarchical as Shavelson et al., (1976) suggested. LISREL model testing could be used to investigate the hypothesis that higher order factor(s) might subsume the six content factors of the Piers-Harris scale.

CONCLUSIONS

Self-concept has been studied as an outcome, moderator and causal variable. With the increased use of self-concept variables in educational and psychological research understanding the nature and dimensionality of self-concept is an important consideration. Much of the research on self-concept and the Piers-Harris in particular has been concerned with the dimensionality of the construct. An approach using LISREL to resolve some of the conflicting issues surrounding the proported theoretical dimensions of the Piers-Harris scale reported in the literature, was illustrated in the present paper. In attempting to do so, the study seems to have brought even more issues to the surface such as the fit of the factors within subgroups and the effect of item phrasing on the measurement.

Previous research directed at studying the dimensions of the Piers-Harris have been limited by the methodology employed. Confirmatory factor analysis while having its own limitations is still thought to be superior to either exploratory factor analysis or multitrait-multimethod matrices in its ability to

analyze the theoretical and structural dimensions of a scale (Hocevar & Hocevar, 1982; Marsh & Hocevar, 1983; Marsh & Smith, 1982; Shvelson & Bolus, 1982). Models can be tested using confirmatory factor analysis across groups to determine the theoretical nature and dimensions of the construct being measured and within groups to determine the stability of the theoretical structure for each subgroup simultaneously. Both analyses help to confirm the interpretation of the scores obtained from the measurement whether the measurement is self-concept or some other latent construct.

REFERENCES

Alwin, D. and Jackson, D. (1981) Applications of simultaneous factor analysis to issues of factorial invariance. In D. Jackson and E. Borgetta (Eds.) Factor Analysis and Measurement in Sociological Research: A Mulitdimensional Perspective. Beverly Hills, Ca.: Sage.

Andrich, D. (1983) Diagnosing and accounting for response sets provoked by items of a questionnaire. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal.

Benson, J & Hocevar, D. (1983) The mpact of item phrasing on the validity of attitude scales for elementary school children. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Bentler, P. (1972) The Piers-Harris Children's Self-Concept Scale (The way I feel about myself). In O.K. Buros (Ed.) Seventh Mental Measurements Yearbook, Highland Park, N.J.: Gryphon Press.

Boomsma, A. (1982) The robustness of LISREL against small sample sizes in factor analysis models. In H. Wald & K. Joreskog (Eds.) Systems Under Indirect Observation. New York: ElsvierNorth-Holland.

Carmines, E. & McIver, J. (1981) Analyzing models with unobserved variables: Analysis of covariance structures. In G. Bohrnstedt and E. Borgatta (Eds.), Social Measurement: Current Issues. Beverly Hills, Ca.

Cowan, R, Altman, H. & Pysh, F. (1978) A validity study of selected self-cocept instruments. Measurement and Evaluation in Guidance, 10, 211,220.

Franklin, M., Duley, S., Pousseau,, E. & Sabers, D. Construct validity of the Piers-Harris Children's Self-concept Scale. Educational and Psychological Measurement, 41, 439-443.

Hocevar, S. & Hocevar, D. (1982) Using confirmatory factor analysis in evaluating factorial validity. Paper presented at the Academy of Mangement, New York.

Hoelter, J. (1983) The analysis of covariance structures: Goodness of fit indices. Sociological Methods in Research, 11, 325-344.

Jersild, A. (1952) In Search of Self. New York: Teachers College, Columbia University, Bureau of Publications.

Joreskog, K. & Sorbom, D. (1983) Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Users Guide, Version V and VI (2nd Ed.). Chicago, Ill.: International Educational Services.

Lynch, M. & Chaves, J. (1975) Do self-concept tests test self-concept An evaluation of the validity of items on the Piers-Harris and Coopersmith measures. National Institute of Education, US Department of Health, Education and Welfare. (ERIC Document No. ED 109 174).

Marsh, H. & Hocevar, D. (1983) Confirmatory factor analysis of multitrait-multimethod matrices. Journal of Educational Measurement, 20, 231-248.

Marsh, H. & Smith, I. (1982) Multitrait-multimethod analyses of two self-concept instruments. Journal of Educational Psychology, 74, 430-440.

Marx, R. & Winne, P. (1978) Construct interpretations of three self-concept inventories. American Educational Research Journal, 15, 99-109.

Marx, R. & Winne, P. (1980) Self-concept validation research: some current complexities. Measurement and Evaluation in Guidance, 13, 72-82.

Moran, M., Michael, W. & Dembo, M. (1978) The factorial validity fo three frequently employed self-report measures of self-concept. Educational and Psychological Measurement, 38, 547-563.

Piers, E. (1969) The Piers-Harris Children's Self-Concept Scale. Nashville, Tenn.: Counselor Recordings and Test, Box 6184, Acklen Station.

Piers, E. & Harris, D. (1964) Age and other correlates of self-concept in children. Journal of Educational Psychology, 55, 91-95.

Platten, M. & Williams, L. (1979) A comparative analysis of the factorial structures of two administrations of the Piers-Harris Children's Self-Concept Scale to one group of elementary school children. Educational and Psychological Measurement, 39, 47-477.

Platten, M. & Williams, L (1981) Replication of a test-retest factoorial validity study with the Piers-Harris Children's Self-Concept Scale. Educational and Psychological Measurement, 41, 453-462.

Robinson, J. & Shaver, P. (1973) Measures of Social Psychological Attitudes. (revised ed.) Ann Arbor, Mi.: Institute for Social Research.

Shavelson, R. & Bolus, R. (1982) Self-concept: The interplay of theory and methods. Journal of Educational Psychology, 74, 3-17.

Shavelson, R., Hubner, J. & Stanton, G. (1976) Self-concept: Validation of construct interpretations. Review of Educational Research, 46, 407-441.

Smith, R. & Rogers, C. (1977) Item instability on the Piers-Harris Children's Self-concept Scale for academic under-achievers with high, medium and low self-concepts: Implications for construct validity. Educational and Psychological Measurement, 37, 553-558.

Wheaton, B., Muthen, B., Alwin, D & Summers, G. (1977) Assessing reliability and stability in panel models. In D. Heise (ed.), Sociolgical Methodology 1977. San Francisco, Ca.: Jossey-Bass.

Winne, P., Marx, R. & Taylor, T. (1977) A multitrait-multimethod study of three self-concept inventories. Child Development, 48, 893-901.

Wylie, R. (1974) The self-concept, Volume I: A review of methodological considerations and measurement instruments. Lincoln, Neb.: University of Lincoln Press.